

CSC2001F 2024 Data Structures

Assignment 1

MSHPRI017

This report is based on a program I created which allows users load data from a knowledge base into an array or binary search tree then query the loaded information, while also incorporating new knowledge or updating the existing knowledge base.

Prince Mashava
Mshpri017@myuct.ac.za

Appropriate design and implementation of OOP:

The data read in from textfiles was formatted in the following way, (term, sentence, confidence score) all separated by tabs. So, I created a class called, Kb which would have these three fields as instance variables and can be used to create an object of type, Kb. The datatypes of the term and sentence are String, and the confidence is of type double.

The constructor of the class takes in a line from the textfile which has a term, sentence, and score. From there the line is sliced into its three parts and the instance variables are assigned their appropriate values.

The class, Kb has get methods that return the term, sentence and confidence score of the object. As well as set methods to modify the instance variables.

There is a toString() method that returns the term, sentence and confidence score altogether separated by tabs.

I also created two methods to search for statements. One is called, matchTerm which takes in a String argument and checks if it matches the term of Kb object exactly. And another is called, matchTermPartial which takes in a String argument and checks if it partially matches the name of the Kb object.

I created a similar method called matchSenPartial which checks if an entered sentence matches the description of a Kb object.

There is a method called replace which will takes in a Kb object as an argument and will replace the term, description and confidence score of the current object.

Traditional Array & BST implementation

The user gets to choose if they want to load the database into an array or BST.

If they choose an array, then an array of Kb objects is used. If the file entered by the user exists then each line from the file will be read using a scanner and for each line, a Kb object is created. Before inserting a new object, the array is checked for an object with the same term. If an object with the same term exists, and it has a lower confidence score than the new Kb object then the object in the array is replaced by the new object with a higher confidence score using the replace method. If an object with the same term is not found, then the object is added to the array. To find a term, a for-loop is used to traverse through the array and check for an entered term.

If they choose a BST, then a BST of Kb objects is used. If the file entered by the user exists then each line from the file will be read using a scanner and for each line, a Kb object is created. Before inserting a new object, the BST is checked for an object with the same term. If an object with the same term exists, and it has a lower confidence score than the new Kb object then the object in the array is replaced by the new object with a higher confidence score using the replace method. If an object with the same term is not found, then the object is added to the BST. To find a term, a recursive is used to traverse through the BST and check for an entered term.

Experimental tests

To test Java program for querying and updating the knowledge base, I looked at various test cases to ensure that the program behaves correctly and handles different scenarios effectively. Those include:

Loading the Knowledge Base:

- Test loading a valid knowledge base file

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 1

Enter file name: GenericsKB.txt
Knowledge base loaded successfully.
```

- Test loading an invalid knowledge base file

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 1

Enter file name: knowledge base.txt
File not found
```

Adding New Statements:

- Test adding a new statement for an existing item with a higher confidence score (Array)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: eye bank

Statement found: An eye bank is a bank (Confidence score: 1.0)

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2

Enter the term: eye bank
Enter the statement: this is an upated version of the eye bank description
Enter the confidence score: 1

Statement for term eye bank has been updated.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: eye bank

Statement found: this is an upated version of the eye bank description (Confidence score: 1.0)
```

- Test adding a new statement for an existing item with a higher confidence score (BST)

```

Enter your choice: 2

Enter the term: screen pass
Enter the statement: this is an updated description
Enter the confidence score: 1

Statement for term, screen pass has been updated.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: screen pass

Statement found: this is an updated description (Confidence score: 1.0)

```

- Test adding a new statement for an existing item with a lower confidence score (Array)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: eye bank

Statement found: An eye bank is a bank (Confidence score: 1.0)

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2

Enter the term: eye bank
Enter the statement: fcgsvdhbjfnkmljdjhgcgsvdhbjfnk
Enter the confidence score: 0.75

A statement for, eye bank already exists and cannot be updated.

```

- Test adding a new statement for an existing item with a lower confidence score(BST)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2

Enter the term: white flower
Enter the statement: this is a new statement with lower confidence
Enter the confidence score: 0.1

Statement for term, white flower has not been updated. The new confidence is lower than the existing one.

```

- Test adding a new statement for a new item (Array)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2

Enter the term: new item
Enter the statement: this is a new item being added to the knowledge base
Enter the confidence score: 0.6

Statement for term new item has been added.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: new item

Statement found: this is a new item being added to the knowledge base (Confidence score: 0.6)

```

- Test adding a new statement for a new item (BST)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2

Enter the term: new item
Enter the statement: this is a new item being inserted
Enter the confidence score: 0.8

Statement for term, new item has been added.

```

Search Results:

- Test displaying information for an existing term (Array)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: tobacco

Statement found: Tobacco increases acid production and damages the lining of the stomach. (Confidence score: 0.7157122492790222)

```

- Test displaying information for an existing term (BST)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: true animal

Statement found: True animals are multicellular and have differentiated tissues. (Confidence score: 0.7352986931800842)

```

- Test displaying information for a non-existing term (Array)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: non-existent

Statement not found

```

- Test displaying information for a non-existing term (BST)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Enter the term to search: non-existent

Statement not found
```

- Test searching for a statement by term and sentence(Array)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: gasp
Enter the statement to search for: A gasp is inhalation

The statement was found and has a confidence score of 1.0.
```

- Test searching for a statement by term and sentence(BST)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: phlox
Enter the statement to search for: Phloxes are plants.

The statement was found and has a confidence score of 1.0.
```

- Test searching for a non-existent statement by term and sentence(Array)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: non-existent
Enter the statement to search for: this statement doesn not appear in the knowledge base

Statement not found
```

- Test searching for a non-existent statement by term and sentence(BST)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: vaseline
Enter the statement to search for: product

Statement not found.
```

Error Handling

- Test handling invalid user inputs (e.g., a confidence score that's not in the 0-1 range)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2

Enter the term: new
Enter the statement: gdhfj
Enter the confidence score: 4
That score is invalid. Please try again.
Enter the confidence score: 5
That score is invalid. Please try again.
Enter the confidence score: 6
That score is invalid. Please try again.
Enter the confidence score: 7
That score is invalid. Please try again.
Enter the confidence score: 88
That score is invalid. Please try again.
Enter the confidence score: 0.5

Statement for term, new has been added.
```

- Test handling invalid user inputs (e.g., a data structure that's not either 1 or 2)

```
mshpri017@nightmare:~/DS_Assignment_01$ make run
java -cp bin main
Select a data structure to load to knowledge base into
Choose your data structure:
1. Array
2. Binary Search Tree

Choose your data structure:
4

That option is invalid. Please try again.
```

Creativity:

I created methods in the Kb class that allows users to search for statements using partial keywords. They can enter a word that appears in the term and a word that appears in the description and if the entered term and sentence are part of a statement's term and description, then a result will be returned to them. I also made all the searches case insensitive so that the user can receive results even if they enter the wrong cases to make the program easier to use.

- Test searching for a statement by partial term and sentence(Array)

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: closer
Enter the statement to search for: magazine

The statement was found and has a confidence score of 1.0.
```

- Test searching for a non-existent statement by partial term and sentence(Array)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: non-existent
Enter the statement to search for: does not appear

Statement not found

```

- Test searching for a statement by partial term and sentence(BST)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: mild hypertension
Enter the statement to search for: heart attack

The statement was found and has a confidence score of 0.7111650109291077.

```

- Test searching for a non-existent statement by partial term and sentence(BST)

```

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4

Enter the term: hypertension
Enter the statement to search for: relax

Statement not found.

```

Git usage log

```

mshpri017@nightmare:~/DS_Assignment_01$ git log | (ln=0; while read l; do echo $ln\: $l; ln=$((ln+1));
done) | (head -10; echo ...; tail -10)
0: commit 5bddc65b9ba02f888f1c33e4b3da3f14083f8ba4
1: Author: Prince Mashava <mshpri017@myuct.ac.za>
2: Date: Fri Mar 8 13:39:26 2024 +0000
3:
4: I created a new file called main.java which will allow the user the select a data structure and then from there, either GenericsKbArrayApp or GenericsKbB
STApp will run. I also added data validation to allow the user to enter 1 for array or 2 for BST. Anything else will result in an error message being displa
yed and them being asked to choose a data structure again.
5:
6: commit 9befde785bc7bf75c36f71cb6ab094d401763581
7: Author: Prince Mashava <mshpri017@myuct.ac.za>
8: Date: Fri Mar 8 08:30:13 2024 +0000
9:
...
40: Author: Prince Mashava <mshpri017@myuct.ac.za>
50: Date: Sun Feb 25 17:56:08 2024 +0000
51:
52: I created a feature that allows users to add new statements in the array using a term. If the term exists and the new confidence sscore is greater than
the current one, then the description of the current statement will be overridden
53:
54: commit f19b3f58aa3a30d8408a9492c6bc463de2783a81
55: Author: Prince Mashava <mshpri017@myuct.ac.za>
56: Date: Sun Feb 25 17:50:56 2024 +0000
57:
58: I created the menu which allows users to select an option when the program starts. I created an array to store the statements from the GenericsKb textfi
le.

```